Working Paper 97-25
Statistics and Econometrics Series 12
February 1997

Departamento de Estadística y Econometría
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (341) 624-9849

NONLINEAR COINTEGRATION WITH MIXING ERRORS.

Alvaro Escribano and Santiago Mira*

Abstract

In this paper we consider an extension of the linear concept of cointegration to a nonlinear context. We discuss the advantages and disadvantages of alternatives concepts of I(0) and I(1) based on the concept of α-mixing and study their relationship with the concept of short memory in distribution. Our concept of nonlinear cointegration can be introduced without having to formally characterize the time series properties of the nonlinear transformations of I(1) variables. The nonlinear least squares (NLS) estimator of the cointegrating relationship is studied under alternative assumptions provided that the nonlinear function is Hadamard differentiable. With some Monte Carlo simulation we found that the bias of NLS estimator can either be large or small depending on the type of nonlinearity allowed in the individual series or in the cointegrating function. We conclude that the proposed framework allows interesting extensions of the classical approach, but is not flexible enough to include several interesting nonlinearities.

Keywords:
Nonlinear cointegration, α-mixing, short memory, nonlinear least squares, hadamard differentiable.

*Escribano, Departamento de Estadística y Econometría, Universidad Carlos III de Madrid. C/ Madrid, 128 28903 Madrid. Spain. Ph: 34-1-624.98.54, Fax: 34-1-624.98.49, e-mail: alvaroe@est-econ.uc3m.es; Mira, Departamento de Fundamentos del Análisis Económico, Universidad Europea de Madrid. The authors acknowledge financial support from spanish DGICYT PB95-0298. We gratefully acknowledge the comments of H.J. Bierens, C.W.J. Granger, M.H. Pesaran, and J. Romo.

1 Introduction

In this paper we extend the linear concept of cointegration to a nonlinear context. There are only few papers that consider the simultaneous treatment of nonstationarity, I(1), and nonlinearity, see for example Escribano (1986, 1987), Granger and Hallman (1991a,b) and Granger and Teräsvirta (1993) and Granger (1995). In particular, Granger and Hallman (1991a) consider the following question; if x_t is I(1), is $g(x_t)$ also I(1) ? As can be expected there is no universal answer to this question and it will depend on the properties of the particular function considered. To answer this question they used a definition of I(0) that satisfies Herrndorf's Functional Central Limit Theorem (FCLT) and they checked if the null hypothesis of the existence of a unit root on the nonlinear transformation was rejected by using a Dickey-Fuller test and a rank Dickey-Fuller test statistics by some Monte Carlo simulations. They found that the Dickey-Fuller test was somehow misleading in this context. The result is not surprising since the usual assumptions on the errors or the Dickey-Fuller regression test are violated in a nonlinear context. In another paper, Granger and Hallman (1991b) study in more detail the possibility of having an attractor (nonlinear cointegration) between long memory processes. They used as a concept of I(0) the idea of short memory in mean and as I(1) the notion of long memory in mean. Those concepts are useful to characterize certain nonlinear cointegration relationships but are no useful to obtain estimation and inference results. The reason is clear, there are no associated Laws of Large Numbers (LLN) nor Functional Central Limit Theorems (FCLT) to them.

In this paper, we propose alternative concepts of I(0) and I(1) that would allow us to obtain estimation and inference results. We do that by using the concept of α-mixing to characterize series that are I(0). Furthermore, we propose a concept of nonlinear cointegration that avoids the difficult characterization of nonlinear functions of I(0) and I(1) variables, see Escribano (1986, 1987). Within this context we study, the asymptotic distribution and the small sample properties of the nonlinear least squares (NLS) estimator of the cointegration relationship.

The structure of the paper is the following. In section 2, we discuss the concepts of I(0), I(1) and nonlinear cointegration based on α-mixing sequences. In section 3, we obtain the asymptotic properties of the least squares estimator of the nonlinear cointegrating relationship. In section 4, we study the small sample properties (biases) of the NLS estimator by some Monte Carlo simulations. Section 5, present some nonlinear examples that are not cover by our framework. Finally, section 6 includes some conclusions.

2 Basic Concepts

As we previously discussed, we will not assume that the series follow linear ARMA models. Therefore, the classical definitions of stochastic trends and I(1) are not appropiate. Granger and Teräsvista (1993) and Granger (1995) propose a natural generalization of those concepts to the nonlinear case. Let $F_h(x) = P(x_{t+h} \leq x|I_t)$ be the conditional distribution of x_{t+h} given the information set $I_t = \{x_{t-j} : j \geq 0\}$. They say that a series x_t is "short memory in distribution" (SMD) if

$$\lim_h F_h(x) = \overline{F}(x)$$

i.e. the conditional distribution does not depends on I_t. Therefore,

$$|P(x_{t+h} \in C_1|x_{t-j} \in C_2) - P(x_{t+h} \in C_1)| \to 0 \quad \text{as} \quad h \to \infty$$

for all subsets $C_1, C_2 \in I_t$ such that $P(x_{t-j} \in C_2) \neq 0$. We can interprete that the concept of ϕ-mixing encapsulates the concept of SMD, but since ϕ-mixing implies α-mixing we will concentrate on the concept of α-mixing.

Definition 2.1 (α-Mixing) Let $\{v_t\}$ be a sequence of random variables. Let $\mathcal{F}_s^t \equiv \sigma(v_s, ..., v_t)$ and define the α-mixing coefficients as

$$\alpha_m \equiv \sup_t \sup_{F \in \mathcal{F}_{-\infty}^t, G \in \mathcal{F}_{t+m}^\infty} |\mathbf{P}(G \cap F) - \mathbf{P}(G)\mathbf{P}(F)| \ .$$

It will be said that the sequence $\{v_t\}$ is α-mixing (or strong mixing) if and only if $\alpha_m \to 0$ as $m \to \infty$. The coefficient α_m measures the dependence between events that depend on v_t's separeted by at least m time periods. The α-mixing property allow simultaneously temporal dependence and heterogeneity in the process. If $\alpha_m = O(m^\lambda)$ for all $\lambda < -\varphi_0$, then it will be said that α_m is of size $-\varphi_0$. Since the concept of α-mixing is based on the σ-algebras generated by the sequence of variables, then the concept is invariant under Borel measurable transformations of a finite number of those variables. See, for instance, White (1984). Examples of α-mixing sequences are the following: Let v_t be defined as $v_t = \sum_{i=0}^q \theta_i w_i$ for $0 \leq q \leq \infty$. If w_i is a sequence of independent variables with with zero mean then $q < \infty$ implies that v_t is α-mixing. If w_i is i.i.d. $\mathcal{N}(0,1)$ and $\sum_{i=0}^\infty |\theta_i| < \infty$ then v_t is α-mixing This case includes any gaussian stationary ARMA model. The requirement of gaussianity is important. Examples of nongaussian AR(1) models that are not α-mixing may be constructed.

Under general conditions there exists a LLN, as the following theorem states.

Theorem 2.2 (McLeish) Let $\{v_t\}$ a scalar α-mixing sequence with α_m of size $r/(r-1)$, $r > 1$, and with finite means $E(v_t) \equiv \mu_t$. If for some δ, $0 < \delta \leq r$, we have

$$\sum_{t=1}^\infty \left(E|v_t - \mu_t|^{r+\delta}/t^{r+\delta}\right)^{1/r} < \infty$$

then $T^{-1}\sum_{t=1}^{T}(v_t - \mu_t) \xrightarrow{a.s.} 0$. □

Proof: See Theorem 3.47 of White (1984).

The condition of Theorem 2.1 is essentially a condition of existence of moments of order $(r+\delta)$, see White (1984). Also under general conditions there exists a FCLT which gives the convergence of partial sums of the α-mixing sequences, as establishes the following theorem.

Theorem 2.3 (Herrndorf) Let $\{v_s\}$ be a sequence of random variables and define $S_T = \sum_{t=1}^{T} v_t$, and $V_T(r) = \sum_{t=1}^{[Tr]} v_t$, where $[Tr]$ is the greater integrer smaller than Tr. Then under assumptions

(i) $E(v_t) = 0$, for all t;

(ii) $\sup_t E(|v_t|^{\beta}) < \infty$, for some $\beta > 2$;

(iii) $\sigma^2 = \lim_{T\to\infty} E(T^{-1}(S_T)^2)$, verifies that $0 < \sigma^2 < \infty$; and

(iv) $\{v_t\}$ is α-mixing with α-mixing coefficients α_m satisfaying

$$\sum_{t=1}^{\infty} \alpha_m^{1-2/\beta} < \infty;$$

we have that $T^{-1/2}V_T(\cdot) \xrightarrow{d} \sigma W(\cdot)$, as $T \to \infty$, where $W(\cdot)$ is the SBM in [0,1]. □

Proof: See Herrndorf (1984).

Condition (ii) controls the existence of moments. Condition (iv) controls the temporal dependence of the process. Since β is the same in (ii) and (iv) there exists a trade off between both, see Phillips (1987). Condition (iii) avoids cases such as the following. Let v_t a Gaussian random walk such that Δv_t ($\Delta v_t \equiv (1-L)v_t \equiv v_t - v_{t-1}$) is a non-invertible MA(1). In that case Δv_t and v_t are α-mixing sequences, but v_t does not satisfy (iii). The following definition of I(1) takes this case into account.

Definition 2.4 (I(0) and I(1)) A sequence $\{v_t\}$ is I(0) if it is α-mixing but the sequence $\{y_t\}$ given by $y_t = \sum_{s=1}^{t} v_t$, is not α-mixing. We will say that y_t is I(1).

Note that if y_t is I(1) then Δy_t is I(0). An alternative definition would be to say that $\{v_t\}$ is I(0) if v_t verifies a FCLT, for instance I(0) may be defined by the Herrndorf conditions. Our definition is more general and only for certain proofs we will require some extra technical conditions. An important property of the above definition is that the α-mixing condition can be tested. There are some papers that deal with this problem. Some of the more important are Lo (1991), Kwiatowski, Phillips, Schmidt and Shin (1992) (KPSS), and Stock (1994). In those papers the null hypothesis is exactly the existence of a FCLT. The authors provide some simulations about the size and power of the tests but none of them provides results on the behaviour of the tests under nonlinearly generated α-mixing sequences.

In what follows we will consider only sequences without deterministic components, i.e., $x_t = \tilde{x}_t - \mu_t$, where μ_t is the mean of $\tilde{x}_t$, such that $E(x_t) = 0$.

Definition 2.5 (Cointegration) Let $\{y_t\}$ and $\{x_t\}$ two I(1) sequences. We will say that y_t and x_t are cointegrated with cointegrating function $g(\cdot, \cdot; \gamma_1^*)$, if $g(y_t, x_t; \gamma_1^*)$ is α-mixing and $g(y_t, x_t; \gamma_1)$ is not α-mixing for $\gamma_1 \neq \gamma_1^*$.

Some comments are worth mention. First, note that we define $g(y_t, x_t; \gamma_1)$ as "not α-mixing" for $\gamma_1 \neq \gamma_1^*$, but we do not specify if $g(y_t, x_t; \gamma_1)$ is I(1). That definition would be inaccurate in the linear case because in that case $g(y_t, x_t; \gamma_1)$ could be I(-1). In this case, however, if $g(y_t, x_t; \gamma_1)$ is not α-mixing, then the dependence has to be stronger, and not weaker.

Second, note that the restriction imposed by the α-mixing condition on the sequence $\{g_t\} = \{g(y_t, z_t; \gamma_1^*)\}$ implies the existence of restrictions on the mean of $\{g_t\}$, but also on every other moment of the sequence.

Third, note that the cointegration function is not unique since any measurable function of an α-mixing sequence is α-mixing. Therefore we will consider the functions $f : \Re^2 \rightarrow \Re$ divided into equivalence classes such that two functions f_1 y f_2 are in the same class if there exists a function $g : \Re \rightarrow \Re$ such that $f_1 = g \circ f_2$. The study will be restricted to one function of each class.

Fourth, note that with this definition new linear cointegration relations appear that were not allowed within the classical cointegration concept, because the dynamics of the variables are not necessarily represented as ARMA models.

Finally, we suppose that the cointegration functions are measurable functions with respect to the appropiate σ-field.

Some extra conditions are implicitly impossed on the cointegration relation in order to avoid certain non-sense cointegration. The following examples specify the relations that are not considered as cointegrating relations. (1) $g(y_t, x_t; \gamma_1) = h(y_t; \gamma_1)$ is a function of only one variable; (2) g is such that for any two variables y_t, x_t of some family of I(1) variables, $g(y_t, x_t; \gamma_1)$ is always I(0) for any value of γ_1, i.e. g gives always cointegration.

The second example tries to avoid "too restrictive" functions. Granger and Hallman (1991b) give the following case. If x_t is a Gaussian random walk, then $\sin(x_t)$ has properties of "short memory". Functions such as $g(y_t, x_t, \gamma_1) = \cos(y_t + \gamma_1 x_t)$, or $g(y_t, x_t, \gamma_1) = \sin(\gamma_1(y_t x_t))$, are therefore "too restrictive" if they always produce cointegration.

It is of interest to consider the "stability" of the definition I(1) for instantaneous transformations. This is due to the fact that the α-mixing property is preserved for such transformations. The following Lemma formalizes a result suggested by Granger and Hallman (1991a).

Lemma 2.6 Suppose four I(1) series given by $\{y_t\}$, $\{\tilde{y}_t\}$, $\{x_t\}$, and $\{\tilde{x}_t\}$, which are related $\tilde{y}_t = f_y(y_t)$, and $\tilde{x}_t = f_x(x_t)$ for invertible transformations $f_y(\cdot)$ and $f_x(\cdot)$.

(a) If there exists a cointegrating function $g_R(\cdot,\cdot)$ for the x_t and y_t series then exists a cointegrating function $g_T(\cdot,\cdot)$ for the $f_x(x_t)$ and $f_y(y_t)$ series.

(b) Conversely, if there exists a cointegrating function $g_T(\cdot,\cdot)$ for the transformed series $\tilde{y}_t$ and $\tilde{x}_t$, then there exists a cointegrating function $g_R(\cdot,\cdot)$ for the series y_t and x_t. $\square$

Proof: See Appendix A.

The invertibility condition of f_x and f_y is not necessary if we impose other restrictions. For instance if we know that $x_t > 0$ then we may consider that $x_t^2 = \tilde{x}_t$ is invertible. Lemma 2.6 have important implications. For example if two I(1) variables x_t and y_t are nonlinear cointegrated then $\log(x_t)$ and $\log(y_t)$ are also nonlinear cointegrated.

Some restrictions on the cointegration function can be stated. Consider the following definition.

Definition 2.7 Given a function $F : \Re^p \rightarrow \Re^q$ s.t. $F(x) = y$ for vectors $x = [x_1, ..., x_p]$ and $y = [y_1, ..., y_q]$, we will say that F is partially invertible if there exists at least one $i \in \{1, ..., p\}$ and one $g_i : \Re^q \rightarrow \Re$ such that $x_i = g_i(y)$.

Note that if F is invertible in the usual sense (i.e. each x_i can be recoved from y) then it is partially invertible. Conversely, if F is not partially invertible, then it is not invertible.

Now suppose we have a cointegration function $K(\cdot)$ such that it transforms an I(1) $(n \times 1)$ random vector x_t into an I(0) $(n \times 1)$ random vector y_t,

$$\begin{aligned} K : \Re^n &\rightarrow \Re^n \\ x_t &\rightarrow K(x_t) = y_t \end{aligned}$$

s.t. $y_t' = K(x_t)' = [K_1(x_t), ..., K_n(x_t)]$ where $K_i : \Re^n \rightarrow \Re$. In this case $K(\cdot)$ is not partially invertible, otherwise we can obtain x_{it} as a function of α-mixing variables y_t, contradicting the initial assumption. See Mira (1996) for more discusions about this issue.

As an aplication of the above statement consider a vector nonlinear error correction (NEC) representation given by

$$\Delta X_t = \Psi_1 \Delta X_{t-1} + H(X_{t-1}) + \varepsilon_t$$

where $H(X_{t-1}) = H(X_{t-1}, \Gamma)$ for some parameter vector Γ. In this case the function $H(X_{t-1})$ must be not partially invertible.

3 Estimation of the Cointegrating Relation

In this section we study the problem of estimation of the cointegrating parameters when the cointegration relationship is nonlinear.

3.1 Some Tools

We will introduce some tools from functional analysis. Let $(\mathcal{F}_1, \|\cdot\|_1)$ and $(\mathcal{F}_2, \|\cdot\|_2)$ be normed spaces, and let $\Psi : \mathcal{F}_1 \to \mathcal{F}_2$ be a functional. We will say that Ψ is differentiable at the point $F \in \mathcal{F}_1$ with respect to a collection of subsets $\mathcal{S}$ of $\mathcal{F}_1$ if there exists a linear continuous map $D\Psi(F;\cdot) : \mathcal{F}_1 \to \mathcal{F}_2$ (which we will call the differential of Ψ at F) such that for G in some neighbourhood of zero,

$$\Psi(F+G) = \Psi(F) + D\Psi(F;G) + R_\Psi(F;G)$$

where the remainder R_Ψ satisfies

$$\lim_{t\to 0} \frac{R_\Psi(F;tG)}{t} = 0$$

uniformly in $G \in S$ for every $S \in \mathcal{S}$. Special choices for $\mathcal{S}$ give the most interesting differentials. If $\mathcal{S}$ is the family of all singletons of $\mathcal{F}_1$ then $D\Psi(F;G)$ is the Gâteaux differential. If $\mathcal{S}$ is the family of all compact subsets of $\mathcal{F}_1$ then $D\Psi(F;\cdot)$ is the Hadamard differential. If $\mathcal{S}$ is the family of all bounded subsets of $\mathcal{F}_1$ then $D\Psi(F;\cdot)$ is the Fréchet differential. Clearly Fréchet differentiability implies Hadamard differentiability, which in turn implies Gâteaux differentiability. In relation with the former definition we have the following theorem, which is a functional version of the well known delta-method theorem.

Theorem 3.1 Suppose $\Psi : \mathcal{F}_1 \to \mathcal{F}_2$ is Hadamard differentiable at $F \in \mathcal{F}_1$ with differential $D\Psi(F;\cdot)$ and that $\{X_T\}_{T=1}^{\infty}$ is a sequence of random elements in $\mathcal{F}_1$ that satisfies:

(i) $T^{-1/2}X_T \xrightarrow{d} X$ in $\mathcal{F}_1$ as $T \to \infty$; and

(ii) the sequence $\{T^{-1/2}X_T\}_{T=1}^{\infty}$ is tight in $\mathcal{F}_1$;

then $T^{-1/2}\Psi(X_T) \xrightarrow{d} D\Psi(0;X)$ in $\mathcal{F}_2$ as $T \to \infty$. □

Proof: The proof is essentially the same as in Heesterman and Gill (1992) with only few changes.

In our case the spaces $(\mathcal{F}_1, \|\cdot\|_1)$ and $(\mathcal{F}_2, \|\cdot\|_2)$ are $(D[0,1]^2, \|\cdot\|_B^2)$, for $D[0,1]$ the space of right continuous with left limits functions (cadlag functions); $\|\cdot\|_B$ is the norm defined by the Skorohod distance modified as in Billingsley (1984), Section 14; and $D[0,1]^2$ and $\|\cdot\|_B^2$ are the double products. Each element X_T, is a function $X_T(\cdot) : [0,1] \to \Re$, with $X_T(r)$ equal to the partial sumation $\sum_{t=1}^{[Tr]} \xi_t$, with $\{\xi\}_{s=1}^{\infty}$ an α-mixing sequence. Therefore, the operator $\Psi(\cdot)$ is given by $\Psi(X_T) = \Psi(\sum_{t=1}^{[Tr]} \xi_t)$. The element X is $W(\cdot)$, the Standard Brownian Motion.

From Theorem 2.2 we have that $T^{-1/2}X_T(\cdot) \xrightarrow{d} \sigma W(\cdot)$, and $\{T^{-1/2}X_T(\cdot)\}$ is tight in $D[0,1]$ (see for instance Herrndorf (1984)), then we have also that

$$T^{-1/2}\Psi(X_T(\cdot)) \xrightarrow{d} D\Psi(0;\sigma W(\cdot))$$

where $D\Psi(0;\sigma W(\cdot))$ is the Haddamar differential of $\Psi(\cdot)$ at zero in the direction of $\sigma W(\cdot)$.

3.2 Estimation of the Cointegrating Parameters

The cointegrating function states that $g_t^* = g_t(\gamma^*) = g(y_t, z_t, \gamma^*)$ is α-mixing and that $g_t = g_t(\gamma) = g(y_t, z_t, \gamma)$ is not α-mixing for $\gamma \neq \gamma^*$. Note that, as in the linear case, under some conditions on g_t^* we have

$$\left(\frac{1}{T}\sum_{t=1}^{T}(g_t^*)^2 - \frac{1}{T}\sum_{t=1}^{T}E(g_t^*)^2\right) \xrightarrow{p} 0 \quad .$$

Therefore to ensure that a nonlinear least squares estimate provides a consistent estimation of γ^* we have to ensure that $\frac{1}{T}\sum_{t=1}^{T}(g_t)^2 \to \infty$ for $g_t \neq g_t^*$. Recall that $y_t = \sum_{s=1}^{t}\eta_s$, and $z_t = \sum_{s=1}^{t}\varepsilon_s$, then the following assumption states a relation between the function $g(\sum_{s=1}^{t}\eta_s, \sum_{s=1}^{t}\varepsilon_s, \gamma)$ and some function $\Phi(\sum_{s=1}^{t}\phi_s, \sum_{s=1}^{t}\delta_s)$ of some α-mixing sequences $\{\phi_s\}$ and $\{\delta_s\}$. Clearly, in general, these sequences will be some elemental transformation of the sequences $y_t = \sum_{s=1}^{t}\eta_s$, and $z_t = \sum_{s=1}^{t}\varepsilon_s$.

Assumption 3.2 (a) There exist a transformation $\Phi(\cdot)$, which is Hadamard differentiable such that the cointegration relation $g(y_t, z_t, \gamma)$ can be writen as $\Phi(\sum_{s=1}^{t}\phi_s, \sum_{s=1}^{t}\delta_s)$ for some strong mixing sequences $\{\phi_s\}$ and $\{\delta_s\}$;

(b) The sequence $\{g_t^*\}$ verifies $T^{-1}\sum_{t=1}^{T}E(g_t^*)^2 \xrightarrow{p} \mu$ as $T \to \infty$.

(c) The α-mixing sequences $\{\phi_s\}$, $\{\delta_s\}$, and $\{g_t^*\}$ satisfy the conditions of Theorem 2.3.

Note that assuming $g(y_t, z_t, \gamma) = \Phi(\sum_{s=1}^{t}\phi_s, \sum_{s=1}^{t}\delta_s)$ is not too restrictive. On the other hand the Haddamard differentiability is a requirement on the smoothness of the (possibly) nonlinear transformation. Clearly the linear case is included as a particular case of the following result.

Theorem 3.3 Under Assumption 3.2, the NLS estimator γ^T which minimizes $\sum_{t=1}^T g(y_t, z_t; \gamma)^2$ provides a consistent estimator of the parameter γ^*. □

Proof: See Appendix A.

In the context of linear cointegrating relationships we know that if the $(n \times 1)$ variable X_t is I(1) and the linear combination $\gamma' X_t$ is I(0), then the OLS estimator $\hat{\gamma}$ of γ is obtained by

$$\hat{\gamma} \in \arg\min_{\gamma \in \Gamma} \sum_{t=1}^T (\gamma' X_t)^2$$

where the restriction Γ is a normalization of the cointegrating vector, such that the linear space generated by the restricted vector $\gamma \in \Gamma$ has to be the same as the space generated by the true γ. The restriction given by $\Gamma = \{\gamma : \gamma = [1, \tilde{\gamma}_2, ..., \tilde{\gamma}_n]'\}$ verifies the required condition and allows us to obtain the estimation by OLS. In the case of r linear cointegrating relationships many possible restrictions are allowed.

In the nonlinear case if the cointegration is given by $f(X_t; \gamma)$ where $f(\cdot, \gamma) : \Re^n \rightarrow \Re$ then the estimation is given by

$$\hat{\gamma} \in \arg\min_{\gamma} \sum_{t=1}^T (f(X_t; \gamma))^2.$$

In this case if $f(X_t, \gamma)$ is α-mixing then also $h(f(X_t, \gamma))$ is α-mixing for Borel measurable $h(\cdot)$ functions. New problems arise related to, but different from, those obtained in the linear case. First, the function $h(f(X_t, \gamma))^2$ may be a function with a maximum around γ^* and then when we find $\min_\gamma \sum_{t=1}^T h(f(X_t, \gamma))^2$ the objective function may be almost flat around the true value γ^* and then the algorithm provides an estimated value quite different from the true value. With an infinite sample the problem vanishes but not with finite samples. With finite samples the normalization proposed is the minimization of $h(f(X_t, \gamma))^2$ for some $h(\cdot)$ which may depend on $f(\cdot)$. See section 4.3 for an example. Second, as in the linear case the function $h(\cdot)$ may depend on a set of parameters γ_2 such that $h(\cdot, \gamma_2^*) = 0$ and then we have an identification problem. For instance, in the linear case the problem is $\min_{\alpha, \beta} \sum_{t=1}^T (\alpha(y_t + \beta x_t))^2$, whose minimum is at $\alpha = 0$.

3.3 Asymptotic Distribution of the Estimator

For the nonlinear case the estimator γ^T of the parameter γ is given by the NLS algorithm. In this case the objective function is

$$\min_{\gamma} \frac{1}{2}\sum_{t=1}^{T} g_t(\gamma)^2 \equiv \min_{\gamma} \frac{1}{2} G(\gamma)'G(\gamma) \equiv \min_{\gamma} Q(\gamma)$$

where the vector $G(\gamma)$ is given by $G(\gamma) = [g_1(\gamma), ..., g_T(\gamma)]'$. If we assume that $\frac{dQ}{d\gamma}(\gamma^T) = 0$, for the estimated value γ^T, then applying a first order Taylor expansion around γ^* we get,

$$0 = \frac{dQ}{d\gamma}(\gamma^T) \approx \frac{dQ}{d\gamma}(\gamma^*) + (\gamma^T - \gamma^*)' \frac{d^2Q}{d\gamma d\gamma'}(\gamma^*) \quad [3.1]$$

therefore

$$(\gamma^T)' \approx (\gamma^*)' - \Big(\frac{dQ}{d\gamma}(\gamma^*)\Big)\Big(\frac{d^2Q}{d\gamma d\gamma'}(\gamma^*)\Big)^{-1}.$$

The following assumption will help us to deal with the nonlinear function $G(\gamma)$.

Assumption 3.4 The function $Q(\gamma)$ verifies $(\frac{d^2Q}{d\gamma d\gamma'}(\gamma^*)) = (\frac{dG}{d\gamma}(\gamma^0)'\frac{dG}{d\gamma}(\gamma^0))$

This assumption is in fact what we use in practice (in each step i) when employing the Gauss-Newton iterative algorithm to obtain the minimum $\min_\gamma Q(\gamma)$. Therefore it is not a restrictive assumption unless we explicitely employ Newton-Raphson as a minimization algorithm. Assuming that this matrix is invertible in γ^* we obtain the relation

$$(\gamma^* - \gamma^T)' \approx \Big(G(\gamma^*)'\frac{dG}{d\gamma}(\gamma^*)\Big)\Big(\frac{dG}{d\gamma}(\gamma^*)'\frac{dG}{d\gamma}(\gamma^*)\Big)^{-1}.$$

Since γ^T is consistent for γ^*, the approximation in [3.1] becomes a equality in the limit, and therefore the asymptotic distribution is given by

$$\begin{aligned} \lim_{T\to\infty} T(\gamma^* - \gamma^T)' &= \lim_{T\to\infty} \Big(T^{-1}G(\gamma^*)'\frac{dG}{d\gamma}(\gamma^*)\Big)\Big(T^{-2}\frac{dG}{d\gamma}(\gamma^*)'\frac{dG}{d\gamma}(\gamma^*)\Big)^{-1} \quad [3.2] \\ &= \lim_{T\to\infty} (T^{-1}V'X)(T^{-2}X'X)^{-1} \end{aligned}$$

for $V = G(\gamma^*)$ and $X = \frac{dG}{d\gamma}(\gamma^*)$. The following theorem gives the convergence to the Standard Brownian Motion in the vectorial case and will be useful later on.

Theorem 3.5 (Phillips and Durlauf) Let $\{x_s\}$ be a sequence of $(k \times 1)$ vectors and let $X_T(r) = \sum_{t=1}^{[Tr]} x_t$, and define $S_T = \sum_{t=1}^{T} x_t = X_T(1)$, then if

(i) $E(x_t) = 0$ for all t;

(ii) $E(T^{-1}S_TS_T') \rightarrow \Sigma$, a positive definite matrix, as $T \rightarrow \infty$ and $E(T^{-1}(S_{K+T}-S_K)(S_{K+T}-S_K)') \rightarrow \Sigma$ as $\min\{K,T\} \rightarrow \infty$.

(iii) $\{x_{it}^2\}$ is uniformly integrable for all $i = 1, ..., k$;

(iv) $\sup_t E(|x_{it}|^\beta) < \infty$ for some $2 \leq \beta < \infty$ and all $i = 1, ..., k$;

(v) $\beta > 2$ and α_m is of size $-\beta/(\beta - 2)$;

then for $W(r)$ the k-dimensional Standard Brownian Motion, and for the decomposition

$$\begin{aligned}
\Sigma_0 &= \lim_{T\rightarrow\infty} T^{-1}\sum_{t=1}^{T} E(x_t x_t') \\
\Sigma_1 &= \lim_{T\rightarrow\infty} T^{-1}\sum_{t=1}^{T}\sum_{j=1}^{t-1} E(x_j x_t') \\
\Sigma &= \lim_{T\rightarrow\infty} E(T^{-1}S_TS_T') = \Sigma_0 + \Sigma_1 + \Sigma_1'
\end{aligned}$$

we have the following results as $T \rightarrow \infty$.

(a) $T^{-1/2}X_T(r) \xrightarrow{d} \Sigma^{-1/2}W(r) \equiv B(r)$;

(b) $T^{-2}\sum_{t=1}^{T} S_t(r)S_t(r)' \xrightarrow{d} \int_0^1 B(r)B(r)'dr$;

(c) $T^{-1}\sum_{t=1}^{T} S_{t-1}x_t' \xrightarrow{d} \int_0^1 B(r)dB(r)' + \Sigma_1$;

(d) $T^{-3/2}\sum_{t=1}^{T} S_t \xrightarrow{d} \int_0^1 B(r)dr$.

for $W(\cdot)$ the SBM k-dimensional. □

Proof: See Lemma 3.1 in Phillips and Durlauf (1986).

Note that in this case $X_T \in D[0.1]^k$, the product metric space of all cadlag real valued functions on $[0.1]$. In this case the definition of α-mixing has to be extended appropiately to the n-dimensional space. The results (a)-(d) hold for the scalar case under assumptions of Theorem 2.2. Write the matrix X as

$$X = \begin{pmatrix} d_1^1 & d_1^2 & \cdots & d_1^k \\ \vdots & \vdots & \ddots & \vdots \\ d_T^1 & d_T^2 & \cdots & d_T^k \end{pmatrix} \equiv \begin{bmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_T \end{bmatrix}$$

where $d_t^j = \frac{\partial g_t}{\partial \gamma_j}$, then the $(k \times k)$ matrix $(X'X)$ can be writen as $\sum_{t=1}^T \mathbf{x}_t'\mathbf{x}_t$, for a $(1 \times k)$ vector $\mathbf{x}_t$. Analogously the $(1 \times k)$ vector $(V'X)$ can be written as $\sum_{t=1}^T g_t^* \mathbf{x}_t$. Let us suppose the following assumption

Assumption 3.6 The derivative $\mathbf{x}_t$ can be written as $\mathbf{x}_t = \sum_{s=1}^t \mathbf{l}_s$ for $\{\mathbf{l}_s\}$ an α-mixing vector sequence, with $\mathbf{l}_s = [l_{1s}, ..., l_{ks}]'$.

Therefore for each $j = 1, ..., k$ we have $\frac{\partial g_t}{\partial \gamma_j} = \sum_{s=1}^t \mathbf{l}_{js}$. It is clear that this assumption holds in the linear case $g_t = \gamma_1 y_t + \gamma_2 x_t$. In the nonlinear case the restriction is restrictive but necessary. Consider an example of a nonlinear cointegrating function given by $g_t = (y_t - \gamma_1)(x_t - \gamma_2)$. In this case $\frac{\partial g_t}{\partial \gamma_1} = -x_t + \gamma_2$ and $\frac{\partial g_t}{\partial \gamma_2} = -y_t + \gamma_1$ and the assumption is satisfied. Now we have the following theorem.

Theorem 3.7 Under Assumptions 3.4 and 3.6 and if the vector $[g_{t-1}^*, \mathbf{l}_t']'$ verify the assumptions of Theorem 3.5, then the asymptotic distribution of the estimator γ^T is given by

$$\begin{aligned}
\lim_{T\to\infty} T(\gamma^T - \gamma^*)' &= \lim_{T\to\infty} (T^{-1}V'X)(T^{-2}X'X)^{-1} \\
&= \lim_{T\to\infty} (T^{-1}\sum_{t=1}^T g_t^*\mathbf{x}_t)(\sum_{t=1}^T \mathbf{x}_t'\mathbf{x}_t)^{-1} \\
&\xrightarrow{d} \left(\int_0^1 \mathbf{B}_2(r)dB_1(r) + \Sigma_{12}\right)\left(\int_0^1 \mathbf{B}_2(r)\mathbf{B}_2(r)'dr\right)^{-1}
\end{aligned}$$

where the parameters are given as follows. $\mathbf{h}_t' = [g_{t-1}^*, \mathbf{l}_t']$ and $\mathbf{k}_t = [\sum_{s=1}^t g_{s-1}^*, \mathbf{x}_t]$; $T^{-2}\sum_{t=1}^T \mathbf{k}_t\mathbf{k}_t' \xrightarrow{d} \int_0^1 \mathbf{B}(r)\mathbf{B}(r)'dr$ and $T^{-1}\sum_{t=1}^T \mathbf{k}_{t-1}\mathbf{h}_t' \xrightarrow{d} \int_0^1 \mathbf{B}(r)d\mathbf{B}(r)' + \Sigma_1$; where $\mathbf{B}(r) = [B_1(r), \mathbf{B}_2(r)']'$ and

$$\Sigma_1 = \begin{pmatrix} \sigma_1 & \Sigma_{12}' \\ \Sigma_{12} & \Sigma_{22} \end{pmatrix}$$

□

Proof: See Appendix B.

Notice that the former theorem ensures the superconsistency of the NLS estimator of the cointegration relationship.

4 Bias in the Estimation of the Cointegrating Parameters

We consider two α-mixing series (in this exercise we will only worry about the α-mixing condition) $\{\eta_s\}$ and $\{\varepsilon_s\}$ and two series $\{y_t\}$ and $\{x_t\}$ given by $y_t = \sum_{s=1}^t \eta_s$ and $x_t =$

$\sum_{s=1}^{t} \varepsilon_s$, such that there exists a function $g(\cdot,\cdot,\gamma^*)$ such that $g(y_t, x_t, \gamma^*)$ is α-mixing. As a by-product, when the function $g(\cdot,\cdot,\gamma^*)$ is linear this approach allows linear cointegration relations that were not allowed in the classical cointegration approach. Within the usual linear framework, a cointegration relation given by $x_t + \alpha y_t$ implies that both x_t and y_t follow ARMA models. With the approach introduced here, those variables may follow any linear or nonlinear models. This section studies the type of biases that appear in the estimation of linear and nonlinear cointegrating relations with I(1) variables on α-mixing sequences.

4.1 Model 1

This case studies the bias that appear when the cointegration is linear and the series are nonlinear transformations of i.i.d. $\mathcal{N}(0,1)$ series. Later on we will allow some temporal dependence on the equilibrium errors. Let us define $\eta_s = v_s + \phi^* v_{s-1} + (m_s - m_{s-1})$ and $\varepsilon_s = v_{s-1} + (n_s - n_{s-1})$, where m_s, n_s and a_s are i.i.d. $\mathcal{N}(0,1)$ and where $v_s = \text{sign}(-a_s)(\beta_2 - \exp(\text{sign}(-a_s)\beta_1 a_s)$, for model 1.1, and v_s is $a_s^3/(a_s^2+1)$ for model 1.2. In this case the cointegration parameter is $\gamma^* = \phi^* + 1$ since

$$
\begin{aligned}
y_t - \gamma^* x_t &= \sum_{s=1}^{t}(v_s + \phi^* v_{s-1}) - \gamma^* \sum_{s=1}^{t} v_{s-1} \\
&= \sum_{s=1}^{t}(v_s + (\phi^* - \gamma^*) v_{s-1}) \\
&= \sum_{s=1}^{t}(v_s - v_{s-1}) = v_t + v_0.
\end{aligned}
$$

The values are $\phi_1^* = 1$, $\beta_1 = 0.8$ and $\beta_2 = 0.5$.

To analyze the behaviour of the estimators we generate N=1000 samples of sizes T=100, T=200 y T=1000, (with the initial 100 data discarded) and we estimate the values γ^T. The following table presents the bias (estimated as the mean $\bar{\gamma}^T = \frac{1}{N}\sum_{i=1}^{N}(\gamma_i^T - \gamma^*)$) and the standard deviation (given by $\sqrt{\frac{1}{N}\sum_{i=1}^{N}(\gamma_i^T - \bar{\gamma}^T)^2}$).

Comparing model 1.1 and model 1.2 we see that the nonlinearity of the equilibrium errors affects the OLS estimation of the cointegrating parameters. In model 1.1 when T is smaller or equal to 500 the bias is large relative to the value of the parameter. For T=1000 the bias if about 10% of the value of the parameter. However, the bias obtained in the OLS estimates in model 1.2 is smaller for size T=100 and even smaller for larger sizes.

Table 1	T=100	T=500	T=1000
Model 1.1	1.1339	0.4936	0.2865
	(0.3599)	(0.2607)	(0.1870)
Model 1.2	0.3768	0.0932	0.0500
	(0.2400)	(0.0708)	(0.0388)

4.2 Model 2

In this case we study the bias that appears when the cointegration relation is linear but the series are nonlinear transformations of ARMA series. Consider $\{v_t\}$ and $\{a_t\}$ as series i.i.d. $N(0,1)$ and define $w_t = \delta w_{t-1} + v_t$, $\rho_t = \log(1 + (0.1)w_t)$ and $\phi_t = a_t - a_{t-1}$. Now define ε_t and η_t as $\eta_t = \pi\rho_t$ and $\varepsilon_t = \rho_t + \lambda\phi_t$. Now, y_t and x_t are generated by the acumulation of η_t and ε_t respectively. If we take $y_t - \gamma^* x_t$ as the cointegrating relation, the equilibrium errors are

$$\begin{aligned} y_t - \gamma^* x_t &= \sum_{s=1}^{t} \left(\eta_s - \gamma^* \varepsilon_s\right) \\ &= \sum_{s=1}^{t} \left(\pi\rho_s - \gamma^*\rho_s - \gamma^*\lambda\phi_s\right) \end{aligned}$$

which are α-mixing for $\gamma^* = \pi$. The values of $\pi = 0.8$, and $\delta = 0.5$ are maintained in both models. Model 2.1 will have $\lambda = 0.4$ and Model 2.2 will have $\lambda = 0.16$. The following table presents the bias of γ^T in the same way as we did in table 1.

With this simulation we see that when λ is large (i.e., ϕ_t has more importance in the errors) the bias is larger. In both cases for size T=500 or greater the biases are smaller than 10% of the value of the parameter. and smaller than in Model 1.

Table 2	T=100	T=500	T=1000
Model 2.1	0.2133	0.0521	0.0245
	(0.1205)	(0.0404)	(0.0228)
Model 2.2	0.0506	0.0091	0.0041
	(0.0440)	(0.0086)	(0.0044)

5 Extensions

Here we discusse a specification that is not included in the framework proposed in this paper. Simulations are conducted to compare the behaviour of this model with that of Models 1 and 2. Take the series n_s and a_s generated by an ARMA(1,1) given by $n_s = 0.6n_{s-1} + 0.8e_{t-1} + e_t$ with e_t i.i.d. $\mathcal{N}(0,1)$ and we define $z_t = \sum_{s=1}^{t} a_s$ and $w_t = \gamma_2(v_t - \gamma_1 z_t)$, with v_s i.i.d. $\mathcal{N}(0,1)$. Now define

$$\begin{aligned} x_t &= \exp(\gamma_2\gamma_1(z_t + 100)/100) + \lambda_1 \\ y_t &= \exp((w_t + n_t + 100)/100) + \lambda_2. \end{aligned}$$

Then the relation g_t given by $g_t = (x_t - \lambda_1)(y_t - \lambda_2)$ is a nonlinear cointegration relation since $g_t = \exp((100\gamma_2\gamma_1 + \gamma_2 v_t + n_t + 100)/100)$ is α-mixing. Note that if the values λ_1 and λ_2 were known in advance the relation $\log(x_t - \lambda_1) + \log(y_t - \lambda_2)$ could have been estimated, but in general they are not known. The parameters λ_1 and λ_2 are estimated by γ_1 and γ_2 given by

$$\min_{\gamma_1,\gamma_2,\gamma_3} \sum_{s=1}^{t} ((x_t - \gamma_1)(y_t - \gamma_2) - \gamma_3).$$

In this case the comment made in Section 3.2 applies. Instead of estimating $(x_t - \gamma_1)(y_t - \gamma_2)$ it is better to estimate the modification previously proposed.

The procedure used to minimize the objective function is the function $ms(\cdot)$ of *S-plus*. In this procedure the initial values given for iterations has been: the mean of x for λ_1, the mean of y for λ_2, and the mean of $x \times y$ for λ_3. The values of the parameters are $\gamma_1 = 0.8$, $\gamma_2 = 0.7$, $\lambda_1 = \lambda_2 = 1$. Table 3 presents the results in the same way as Tables 1 and 2. It can be seen that for $T = 100$ the bias are quite large and they decrease slowly. For $T = 1000$ the bias is still around a 25% of the value of the parameter. Therefore, when the cointegration relation is nonlinear the NLS estimator might need very large samples to give small biases.

Table 3	T=100	T=500	T=1000
λ_1	1.4758	0.5345	0.2207
	(0.5500)	(0.4644)	(0.2689)
λ_2	2.3163	0.7929	0.3272
	(0.8506)	(0.6713)	(0.3969)

6 Conclusions

In this paper we have investigated what type of nonlinearities and nonstationarities can be study by NLS based on mixing errors. This is of interest since the mixing concept is preserved under nonlinear transformations. We have proposed concepts of I(0), I(1) and nonlinear cointegration that are useful to study the asymptotic properties of the nonlinear least squares (NLS) estimator of the cointegrating relationship. We consider nonlinear cointegrating functions that are Hadamard differentiable. For those definitions of I(1) and cointegration we give conditions for the consistency of the NLS estimator. However, to get the superconsistency and the asymptotic distribution of the NLS we need to add more structure. Two of the relevant additional assumptions are that the I(0) variables satisfy a FCLT and that the first order derivatives of the nonlinear cointegrating function have to be I(1). Certainly this last condition on the derivative and our definition of I(1) rules out interesting nonlinearities and that extension deserves further research. The small sample properties of the NLS are informative. First, we have seen that even a simple linear cointegrating relationship estimated by OLS can have large biases if the underlying sequences are a-mixing but nonlinear. Theoretically, this situation was included in Phillips and Durlauf(1986) framework but no simulation were done to check its small sample properties. Therefore, the nonlinearity of the residuals can affect or not affect the small sample bias of the OLS cointegrating estimator depending on the type of nonlinearity considered. Within a more general nonlinear context, we could have small sample biases when estimating by NLS. Once again those biases depend on the type of nonlinearity of the residuals, but some biases are large even for sample sizes of 1000 observations. As we have said before some of the concepts and assumptions used in this paper are too restrictive. However, if we relax them we would need to find weaker conditions for a FCLT to hold, maybe with a different concept than α-mixing, and maybe we need to consider nonparametric estimators, instead of NLS, but all these questions are out of the scope of this paper.

A Appendix to Section 2

A.1 Proof of Lemma 2.6

For the first part define $w_t = f_x(x_t)$ y $r_t = f_y(y_t)$. Now, define $g_T(w_t, r_t) = g_R(f_x^{-1}(w_t), f_y^{-1}(r_t))$. Clearly $g_R(f_x^{-1}(w_t), f_y^{-1}(r_t)) = g_R(x_t, y_t)$ and then it is α-mixing.

The second part is more straightforward. Define $g_R(x_t, y_t) = g_T(f_x(x_t), f_y(y_t))$ and the result follows.

Q.E.D.

B Appendix to Section 3

B.1 Proof of Theorem 3.3

We will prove that $T^{-1}\sum_{t=1}^{T} g(y_{t-1}, z_{t-1}, \gamma)^2 \to \infty$. To do that we will use Theorem 3.1. We will write g_{t-1}^2 instead of the expression $g(y_{t-1}, z_{t-1}, \gamma)^2$. Then from the assumptions we can write

$$
\begin{aligned}
T^{-2}\sum_{t=1}^{T} g_{t-1}^2 &= T^{-1}\sum_{t=1}^{T}(\Phi(\sum_{s=1}^{t-1}\phi_s, \sum_{s=1}^{t-1}\delta_s))^2 T^{-1} \\
&= T^{-1}\int_0^1 M_T(r)dr
\end{aligned}
$$

where $M_T(r)$ is given by

$$
M_T(r) = \begin{cases}
0 & \text{for } 0 \leq r < \frac{1}{T} \\
(\Phi(\phi_1, \delta_1))^2 & \text{for } \frac{1}{T} \leq r < \frac{2}{T} \\
\vdots & \vdots \\
(\Phi(\sum_{s=1}^{T-1}\phi_s, \sum_{s=1}^{T-1}\delta_s))^2 & \text{for } \frac{T-1}{T} \leq r < 1 \\
(\Phi(\sum_{s=1}^{T}\phi_s, \sum_{s=1}^{T}\delta_s))^2 & \text{for } r = 1
\end{cases}.
$$

Now we have the following convergences

$$
T^{-1/2}\sum_{s=1}^{[Tr]}\phi_s \xrightarrow{d} \sigma_1 W_1(r)
$$

$$
\begin{aligned}
T^{-1/2}\sum_{s=1}^{[Tr]}\delta_s &\xrightarrow{d} \sigma_2 W_2(r) \\
T^{-1/2}\Phi(\sum_{s=1}^{[Tr]}\phi_s,\sum_{s=1}^{[Tr]}\delta_s) &\xrightarrow{d} D\Phi(0;\sigma_1 W_1(r),\sigma_2 W_2(r)) \\
T^{-1}M_T(r) \equiv T^{-1}(\Phi(\sum_{s=1}^{[Tr]}\phi_s,\sum_{s=1}^{[Tr]}\delta_s))^2 &\xrightarrow{d} (D\Phi(0;\sigma_1 W_1(r),\sigma_2 W_2(r)))^2 \equiv \widetilde{W}(r)^2 \\
\int_0^1 T^{-1}M_T(r)dr &\xrightarrow{d} \int_0^1 \widetilde{W}(r)^2 dr
\end{aligned}
$$

Since $T^{-2}\sum_{t=1}^T g_{t-1}^2 \xrightarrow{d} \int_0^1 \widetilde{W}(r)^2 dr$, then $T^{-1}\sum_{t=1}^T g_t^2 \to \infty$, and the NLS estimator γ^T given by $\min_\gamma Q_T(\gamma)$ where $Q_T(\gamma) = T^{-1}\sum_{s=1}^T g_t(\gamma)^2$, provides a consistent estimation of γ^*.
Q.E.D.

B.2 Proof of Theorem 3.7

Let us define the $((k+1)\times 1)$ vectors

$$
\mathbf{h}_t = \left[\begin{array}{c} g_{t-1}^* \\ \mathbf{l}_t \end{array}\right] = \left[\begin{array}{c} h_{1t} \\ \mathbf{h}_{2t} \end{array}\right] \quad \text{and} \quad \mathbf{k}_t = \sum_{s=1}^{t}\mathbf{h}_s = \left[\begin{array}{c} \sum_{s=1}^t g_{s-1}^* \\ \mathbf{x}_t \end{array}\right] = \left[\begin{array}{c} k_{1t} \\ \mathbf{k}_{2t} \end{array}\right].
$$

If we apply Theorem 3.5 we obtain the following convergences to the $((k+1)\times(k+1))$ matrices

$$
\begin{aligned}
T^{-2}\sum_{t=1}^{T}\mathbf{k}_t\mathbf{k}_t' &\xrightarrow{d} \int_0^1 \mathbf{B}(r)\mathbf{B}(r)'dr \\
T^{-1}\sum_{t=1}^{T}\mathbf{k}_{t-1}\mathbf{h}_t' &\xrightarrow{d} \int_0^1 \mathbf{B}(r)d\mathbf{B}(r)' + \Sigma_1
\end{aligned}
$$

where the $((k+1)\times 1)$ vector $\mathbf{B}(r)$ is given by $\mathbf{B}(r) = [B_1(r), \mathbf{B}_2(r)']'$ and

$$
\Sigma_1 = \left(\begin{array}{cc} \sigma_1 & \Sigma_{12}' \\ \Sigma_{12} & \Sigma_{22} \end{array}\right)
$$

and an analogous decomposition can be made for Σ. Now we have the following convergences

$$
\begin{aligned}
T^{-2}X'X = T^{-2}\sum_{t=1}^{T}\mathbf{x}_t'\mathbf{x}_t = T^{-2}\sum_{t=1}^{T}\mathbf{k}_{2t}'\mathbf{k}_{2t} &\xrightarrow{d} \int_0^1 \mathbf{B}_2(r)'\mathbf{B}_2(r)dr \\
T^{-1}V'X = T^{-1}\sum_{t=1}^{T} g_t^*\mathbf{x}_t = T^{-1}\sum_{t=1}^{T} h_{1,t+1}\mathbf{k}_{2t} &\xrightarrow{d} \int_0^1 \mathbf{B}_2(r)dB_1(r) + \Sigma_{12}
\end{aligned}
$$

and the result follows. Note that $T^{-1}\sum_{t=1}^{T} h_{1,t+1}\mathrm{k}_{2t} = T^{-1}\sum_{t=1}^{T} h_{1t}\mathrm{k}_{2,t-1} + o_p(1)$.
Q.E.D.

References

Billingsley, P., (1984), *Convergence of Probability Measures.* New York: Wiley.

Escribano, A., (1986), *Non-Linear Error-Correction: The Case of Money Demand in the U.K. (1878-1970).* Chapter IV. Ph. D. Dissertation. University of California, San Diego.

Escribano, A., (1987), "Error-Correction Systems: Nonlinear Adjustment to Linear Long-Run Relationships", *CORE Discussion Paper 8730. C.O.R.E..*

Granger, C. W. J., (1995), "Modelling Nonlinear Relationships Between Extended-Memory Variables", *Econometrica, 63 (2), 265-279.*

Granger, C. W. J., and J. Hallman, (1991a), "Nonlinear Transformations of Integrated Time Series", *Journal of Time Series Analysis, 12, 207-224.*

Granger, C. W. J., and J. Hallman, (1991b), "Long Memory Series with Attractors". *Oxford Bulletin of Economics and Statistics, 53, 11-26.*

Granger, C. W. J., and T. Teräsvirta, (1995), *Modelling Nonlinear Economic Relationships.* New York: Oxford University Press.

Heesterman, C. C., and R. D. Gill, (1992), "A Central Limit Theorem for M-estimators by the von Mises Method". *Statistica Neerlandica, 46, 165-177.*

Herrndorf, N., (1984). "A Functional Central Limit Theorem for Weakly Dependent Sequences of Random Variables". *Annals of Probability, 12, 141-153.*

Kwiatkowski, D., P. C. B. Phillips, P. Schmidt, and Y. Shin, (1992), "Testing the Null Hypothesis of Stationarity Against the Alternative of a Unit Root", *Journal of Econometrics. 54. 159-178.*

Lo, A. W., (1991). "Long-Term Memory in Stock Market Prices", *Econometrica, 59, 1279-1313.*

Mira, S., (1996), *Modelos Econométricos Dinámicos No Lineales con Tendencias Estocásticas.* Ph. D. Dissertation. Universidad Carlos III de Madrid. Madrid.

Phillips, P. C. B., (1987), "Time Series Regression with a Unit Root", *Econometrica, 55, 277-301.*

Phillips, P. C. B., and Durlauf, S. N., (1986), "Multiple Time Series Regression with Integrated Processes", *Review of Economic Studies, 473-495.*

Stock, J. H., (1994), "Deciding between I(1) and I(0)", *Journal of Econometrics, 63, 105-131.*

White, H., (1984), *Asymptotic Theory for Econometricians.* New York: Academic Press.